Handling what the other sees: the effects of seeing and being seen on gesture production

Lisette Mol (L.Mol@uvt.nl)

Emiel Krahmer (E.J.Krahmer@uvt.nl)
Tilburg Centre for Cognition and Communication (TiCC), School of Humanities, Tilburg University
P.O. Box 90135, NL-5000 LE Tilburg, The Netherlands

Abstract

Language production is often argued to be adapted to addressees' needs. As an instance of this, speakers produce fewer speech accompanying hand gestures if the speaker and the addressee cannot see each other. Yet there is also empirical evidence that speakers tend to base their language production on their own perspective, rather than their addressee's. Therefore, speakers may gesture differently because they do not see their addressee, rather than because their addressee cannot see them. Can speakers truly apply their knowledge of what their addressee sees to their gesture production? We answered this question by carrying out a production experiment in which visibility between speaker and addressee was manipulated asymmetrically. We found that representational gestures were produced more frequently when speakers could be seen by their addressee, rather than when they could see their addressee, suggesting that speakers indeed apply their knowledge of the addressee's perspective correctly to their gesturing.

Keywords: Gesturing, Audience Design.

Introduction

Language use sometimes requires taking into account what another person can or cannot see. For example, when watching a documentary on Venice with a friend, you might ask your friend "have you ever been there?", where *there* refers to Venice. However, if your friend was in the same room, but working on her computer "have you ever been to Venice?" may be more appropriate. Because you know your friend is not watching the documentary, you may choose a more explicit reference. On the other hand, if you were asked by your friend, "have you ever been there?", while working on your computer, your knowledge of her watching a documentary on Venice may help in arriving at the correct interpretation. Yet would you do so correctly if you happened to be browsing a website on Cologne?

Language production is often argued to be adapted to the needs of addressees (e.g. Grice, 1989). As an instance of this, it is well established that speakers produce fewer speech accompanying hand gestures when interlocutors cannot see each other (Cohen & Harison, 1973). Yet several empirical studies suggest that applying knowledge of what another person can and cannot see is not at all straightforward (e.g. Keysar, Lin, & Barr, 2003; Wardow Lane, Groisman, & Ferreira, 2006). These studies suggest that interlocutors tend to base their language use on their own perspective, rather than that of their conversation partner.

To our knowledge, in studies on hand gestures, visibility has always been manipulated symmetrically. That is, whenever the addressee could not see the speaker, neither was the speaker able to see the addressee. Therefore, these studies cannot reveal whether it is the speaker's own perspective that underlies this reduction in gesture frequency, or whether speakers adapt their language use to their addressee's perspective. In this study we aim to fill this gap, by manipulating visibility asymmetrically. For this we make use of computer-mediated communication. We will therefore also make a comparison of our data in computer-mediated settings to data acquired in similar unmediated settings (Mol, Krahmer, Maes, & Swerts, 2009).

Taking into Account what an Interlocutor sees

Keysar, Lin, and Barr (2003) have shown that people make 'mistakes' in interpreting speech, when deriving the correct interpretation requires applying ones knowledge of what the speaker does not see. In their study, a follower had visual access to an object that was occluded from the director's view. Still, when the (confederate) director's description more closely resembled the hidden object than any of the mutually visible objects, the follower often considered this object as a referent, sometimes even moving it instead of the intended object. This shows that the follower's knowledge of what the director could (not) see was not automatically applied to the interpretation process.

Wardlow Lane, Groisman, and Ferreira (2006) found similar results for reference production. In their study a speaker had private visual access to an object that only differed from the target object in size. Even though the addressee could not see this competing object, speakers often included a contrasting adjective, such as 'small' in their reference to the target object. Surprisingly, they did so even more when instructed to conceal their private information from the addressee. Thus, it seems that speakers have difficulty in applying their knowledge of what their addressee can see to the speech production process as well.

Gesturing out of Sight

The question naturally arises whether knowledge of what another person sees is applied correctly to the production of

co-speech hand gestures. These gestures are spontaneous movements of the hands and arms during speech (e.g. McNeill, 1992). Hand gestures can, amongst other functions, be communicative. For example they can convey meaning (e.g. Beattie & Shovelton, 1999) or emphasize certain parts of speech (e.g. Hadar, 1989; Krahmer & Swerts, 2007). It has been found repeatedly that speakers' gesturing differs depending on whether their addressee can see them or not (e.g. Alibali, Heath, & Myers, 2001; Bavelas, Gerwing, Sutton, & Prevost, 2008; Cohen & Harison, 1973). For example, Alibali et al. asked participants to retell the story of an animated cartoon to an addressee. During half of the narration, an opaque screen separated speaker and addressee, such that no information could be conveyed through hand gestures. They found that speakers gestured less frequently when the screen was in place. This was especially true for *representational gestures*, which depict some of the content a speaker is trying to convey. It thus seems that at least some gesturing is influenced by the speaker's knowledge of what the addressee can and cannot see.

However, in the studies cited above, visibility was always manipulated symmetrically. That is, the addressee could not see the speaker, but neither could the speaker see the addressee. It is thus possible that speakers used their own perspective, and that their gesturing changed as a result of them not seeing the addressee, rather than of them correctly applying their knowledge of what the addressee could see. If so, many other factors may have influenced gesture production, such as perceived attentiveness of the addressee, social fulfillment during the task, general motivation, etc. Indeed, Jacobs and Garnham (2006) found that people gesture less frequently towards an addressee who appears to be less interested. Interest can be conveyed by gaze (Argyle & Cook, 1976), and also by body posture and head nods, which are all absent if visibility is obstructed. It is therefore still unclear whether the reduced frequency of hand gestures when interlocutors cannot see each other is an instance of the correct application of the knowledge the speaker has about the addressee's visual perspective.

Desktop Video-Conferencing

One way to manipulate visibility in an asymmetrical way is by computer-mediated communication. Yet is mediated communication representative of unmediated communication? Brennan and Oheari (1999) found evidence that mediated communication may differ from unmediated communication as a direct result of the differences in affordances between the media, rather than for example because interlocutors become less socially aware when they are not physically copresent. In typing - which is often used in mediated communication - different types of communicative behavior are effortful than in speech. Brennan and Oheari found that especially back-channeling behavior differed between spoken and written dialogue. This in turn may affect interlocutors' perception of each other, rather than them not being physically co-present. Thus, the more affordances mediated communication offers, the more similar it will be to unmediated communication.

Modern video-conferencing tools allow speakers to see and hear each other even though they are in different locations. Isaacs and Tang (2003) observed interactions between technical experts that took place over the phone, through desktop video-conferencing, or face-to-face. They found that the experts used the visual modality in video-conferencing much like they did in face-to-face communication. "Specifically, participants used the visual channel to: express understanding or agreement, forecast responses, enhance verbal descriptions, give purely nonverbal information, express attitudes through posture and facial expression, and manage extended pauses", p. 200. They also list some differences between video-conferencing and face-to-face communication, for example, managing turn-taking, having side conversations, and pointing towards objects in each other's space were more difficult in video-conferencing.

In the video-conferencing we use, interlocutors can communicate through speech as though they are in the same room. The need for turn-taking is minimal, and there are only two interlocutors. Also, our task is not about manipulating the environment, which reduces the factor of not sharing a workspace. We therefore expect that manipulating mutual visibility will have similar effects in our mediated settings as it does in unmediated settings. But more readily than unmediated communication, video-conferencing enables one-way visibility, allowing for example the speaker to see the addressee, but not vice versa. It is thus very suitable for testing whether or not speakers employ an egocentric perspective when they cannot see their addressee.

Present Study

In this study we aim to gain insight into whether people generally employ an egocentric perspective in their language production. We address this question by testing if speakers' knowledge of whether their addressee can see them or not influences their co-speech gesturing. We manipulate visibility asymmetrically. That is, some speakers will be able to see their addressee, but will know that the addressee cannot see them, and some speakers will not be able to see their addressee, but will know that the addressee can see them. If gesturing is based on the speaker's own visual perspective, then gesturing will be more frequent when speakers can see the addressee, regardless of whether the addressee can see them. This could be either because the addressee seems more engaged or more present when visible, or because from the speaker's visual perspective, it seems as though speaker and addressee can see each other. Yet if speakers correctly apply their knowledge of the addressee's visual perspective, then they are expected to

gesture more when the addressee can see them, regardless of whether they can see the addressee. If both of these factors increase gesture production, then gesturing should be most frequent when interlocutors can see each other.

Method

Design

We have used a 2 x 2 between subjects design in which we manipulated whether or not the addressee could see the speaker and whether or not the speaker could see the addressee. In all conditions speaker and addressee could hear each other.

Participants

38 (21 female) native Dutch speakers, all students of Tilburg University, participated in this study as part of their first year curriculum. Two participants were excluded from our analysis (see Coding and Analysis). The remaining 36 participants (20 female) had a mean age of 22, range (18 - 30). The addressee was a female confederate, who was also a student at Tilburg University.

Procedure

The participant and the confederate were received in the lab by the experimenter, who assigned the role of speaker to the participant and the role of addressee to the confederate. Like in the study by Alibali et al. (2001), narrators were asked to retell the story of an animated cartoon (*Canary Row* by Warner Bro's). After reading the instructions participants could ask any remaining questions. (The confederate always posed a question.) The narrator's instructions stated that the addressee had to summarize the narration afterwards and explained that the narrator was videotaped in order to compare the summary to the narration afterwards.

When all was clear the narrator was seated behind a table with a computer screen on it, which in some settings showed a live video-image of the addressee, and in the remaining settings showed the interface of a video-conferencing application (Skype). The screen was connected to a pc, which also had a web cam connected to it. Behind the table stood a tripod, which held the web cam and a digital video camera. On the wall behind the video camera were eight stills from the animated cartoon, one from each episode, as a memory aid for the narrator and to elicit more structured and hence more comparable narrations.

The experimenter took the addressee to another room with a similar setup (but without the stills) and established a connection between the two pc's over the internet, using Skype. Sound and video were both captured by the web cams and sound was played back through speakers. Sound was tested by the narrator and addressee talking to each other and if applicable, the video image was tested by them watching each other. The connection was then suspended temporarily while the narrator was left alone to watch the animated cartoon on a different computer. When the cartoon had finished the experimenter re-established the connection, and seated the narrator behind the camera. The experimenter repeated whether the addressee could see the narrator or not, started the video recording, and left the room.

Figure 1: Left: example of a representational gesture (depicting hitting), Right: example of a non-representational gesture (placing emphasis while referring to a character).

When the narrator was done telling the story, a questionnaire followed, which included questions on how the communicative setting had been experienced, how interested the addressee had appeared, whether any deception was suspected, and finally whether the participant was left or right handed. Meanwhile, the addressee ostensibly wrote a summary on yet another computer in the lab room. None of the participants had suspected any deception. After filling out the questionnaire, they were fully debriefed. All of the participants gave their informed consent for the use of their data, and if applicable for publishing their photographs.

During the narration, the confederate refrained from interrupting, laughing, etc. When necessary, minimal feedback was provided verbally. She always gazed somewhere near the web cam capturing her, independent of whether she could see the speaker.

Coding and Analysis

Video recordings of all narrators were coded using Noldus Observer. For each movement of the hands it was determined whether the movement was a gesture or a self-adaptor. Gestures were labeled as either *representational*, expressing some of the content of the speaker's story, or *non-representational*, placing emphasis or regulating interaction. Figure 1 depicts two examples. In the scene on the left, the speaker imitates a hitting motion while talking about someone hitting. In the scene on the right, the speaker refers to the main character and briefly moves his fingers up and down. In order to normalize for the duration of each speaker's narration, we have used the number of gestures produced per minute as the dependent variable, rather than the total number of gestures produced.

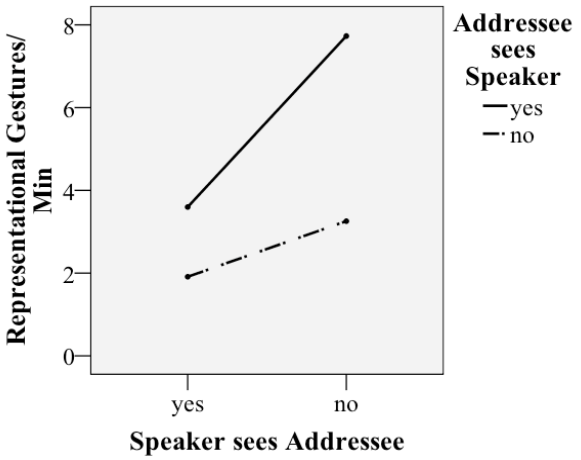

Figure 2: Means of the rate of representational gestures across settings.

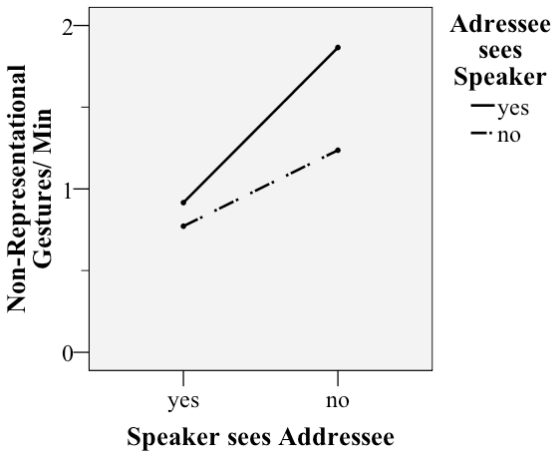

Figure 3: Means of the rate of non-representational gestures across settings.

The analysis was done using ANOVA, with fixed factors *addressee sees speaker* (yes, no) and *speaker sees addressee* (yes, no). Our significance threshold was .05 and we have used partial eta squared as a measure of effect size.

Two participants were excluded from the analysis, because they deviated more than 2 standard deviations from the mean gesture rate in their condition. As a result, there were 9 participants in each condition. Inclusion of these two participants did not affect the significant effects found, but did reduce the significance of the overall model.

Results and Discussion

We did not find an effect of gender or left or right handedness on gesture rate, or on the total duration of the narration. Neither did we find an effect of condition on the duration of the narration.

Effect of the Addressee seeing the Speaker

Figure 2 shows the mean number of representational gestures per minute in each setting. Whether or not the addressee could see the speaker reliably influenced this gesture rate, $F(1, 32) = 4.873$, $p < .05$, $\eta^2 = .13$. When speakers could be seen by the addressee, they produced representational gestures more frequently ($M = 5.7$, $SD = 5.8$) than when they could not be seen ($M = 2.6$, $SD = 3.4$). We found no significant effect of visibility of the speaker on the rate of non-representational gestures ($p = .35$).

Effect of the Speaker seeing the Addressee

The effect of whether the speaker could see the addressee approached significance for the rate of representational gestures, $F(1, 32) = 3.854$, $p = .06$, $\eta^2 = .11$. When speakers could see their addressee, they produced these gestures *less* frequently ($M = 2.8$, $SD = 3.4$) than when they could not see their addressee ($M = 5.5$, $SD = 5.3$). There was no significant interaction between visibility of the speaker and addressee ($p = .33$).

The mean number of non-representational gestures in each condition is depicted in Figure 3. The effect of the speaker seeing the addressee on this gesture rate showed a trend towards significance, $F(1,32) = 2.977$, $p = .09$. Non-representational gestures were produced less frequently when speakers could see their addressee ($M = .84$, $SD = .87$), compared to when they could not ($M = 1.6$, $SD = 1.5$). There was no significant interaction with the addressee seeing the speaker ($p = .56$).

Perceived Interest

Our questionnaire revealed that in the setting in which the speaker could see the addressee but not vice versa, the addressee was perceived as significantly more uninterested than in any of the other conditions, $F(3, 31) = 5.232$, $p < .01$, see Table 1. (Pairwise comparisons were done using the LSD method with a significance threshold of .05.)

Discussion

When the addressee could see the speaker, speakers produced representational hand gestures more frequently than when the addressee could not see them. This was true both when the speaker could see the addressee and when

Table 1: *Means* and *Standard Deviations* of speakers' answer to the statement "The addressee was disinterested" on a 7 point scale, 1 = completely disagree, 7 = strongly agree.

Addressee sees Speaker	Speaker sees Addressee	Mean, *SD* of Perceived disinterest (1 to 7 scale)
Yes	Yes	2.7, 1.0
Yes	No	3.3, 1.3
No	Yes	4.5, 1.2
No	No	2.4, 1.1

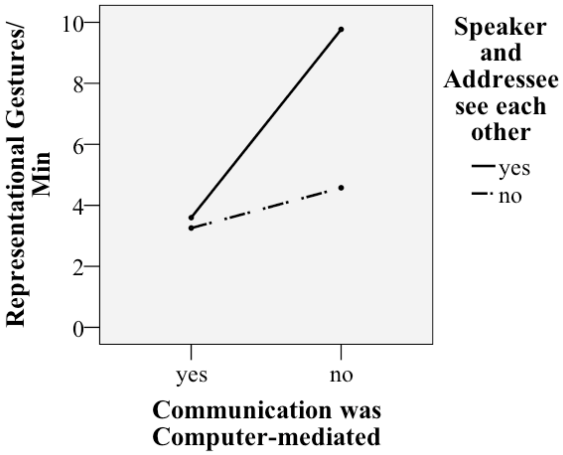

Figure 4: Means of the rate of representational gestures in mediated and unmediated settings.

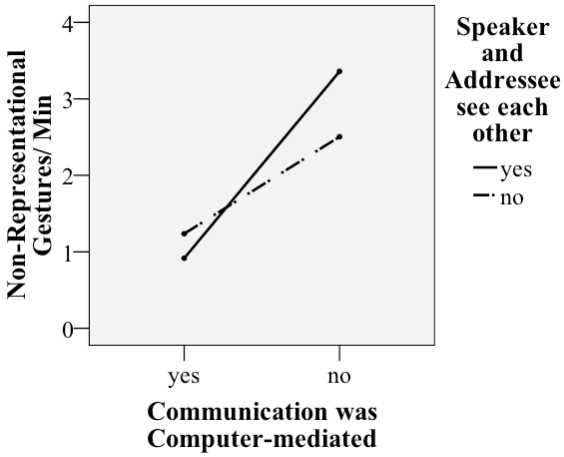

Figure 5: Means of the rate of non-representational gestures in mediated and unmediated settings.

not. We therefore conclude that the knowledge a speaker had about what the addressee could and could not see was incorporated correctly into their hand gesture production.

We found an unexpected effect when speakers could see their addressee. When they saw a live video-image of their addressee, speakers produced representational gestures less frequently and also tended to produce non-representational gestures less frequently than when they did not see their addressee. This would be understandable if the addressee came across as less interested when visual feedback was provided. In the setting in which the addressee could not see the speaker, there was nothing relevant to look at for the addressee. To keep the settings comparable, the addressee therefore always gazed somewhere near the web cam capturing her. This may have been interpreted as lack of interest. The answers to our questionnaire support this hypothesis. In the setting in which the speaker could see the addressee but not vice versa, the addressee was rated as significantly less interested than in all other settings.

Mediated vs. Unmediated Settings

In the study above, we manipulated visibility by means of computer-mediated communication. In an earlier study (Mol et al. 2009), we have manipulated visibility while speaker and addressee were in the same room. The procedure was the same as in the current study, except that the speaker and addressee were in the same room facing each other (N = 10), or in the same room but separated by an opaque screen (N = 9). Given that the affordances in these mediated and unmediated settings are a close match, it is interesting to see whether there still is an effect of computer-mediation. To address this question we compare the mediated settings with mutual visibility and with audio only to their unmediated counterparts. Participants were mostly first year students of Tilburg University and all were native speakers of Dutch. The mean age was 19, range (17 – 21), and 15 out of 19 participants were female.

Effect of Visibility

The gesture rates across settings for representational gestures are depicted in Figure 4. The main effect of visibility on this gesture rate approached significance, $F(1, 33) = 4.1$, $p = .05$. Participants gestured more frequently when they could see each other ($M = 6.8$, $SD = 6.1$) than when they could not ($M = 3.9$, $SD = 2.3$). There was no significant effect of mutual visibility on the rate of non-representational gestures ($p = .65$).

Effect of Mediation

Mediation had a significant main effect on the rate of representational gestures, $F(1, 33) = 7.579$, $p < .01$. The interaction between mutual visibility and mediation showed a trend towards significance, $F(1, 33) = 3.180$, $p = .08$. The difference between the visibility and no visibility condition was larger in the unmediated settings.

Mediation also influenced the rate of non-representational gestures, $F(1, 33) = 10.330$, $p = .01$. Non-representational gestures were produced more frequently in the unmediated settings ($M = 3.0$, $SD = 2.2$), compared to the mediated settings ($M = 1.1$, $SD = 1.1$). There was no significant interaction between the factors ($p = .32$). The gesture rates for non-representational gestures are depicted in Figure 5.

Perceived Interest

The effect of the setting on how disinterested the addressee was perceived showed a trend towards significance, $F(3, 33) = 2.288$, $p = .097$. Table 2 (next page) shows the means and standard deviations for this measure in each setting. Pairwise comparisons with the LSD method showed that addressees were perceived as less interested in the unmediated setting without visibility, compared to the unmediated setting with visibility and the mediated setting without visibility, $p < .05$.

Table 2: *Means* and *Standard Deviations* of speakers' answer to the statement "The addressee was disinterested" on a 7 point scale, 1 = completely disagree, 7 = strongly agree.

Mutual Visibility	Computer-Mediation	Mean, *SD* of Perceived disinterest (1 to 7 scale)
Yes	Yes	2.7, 1.0
Yes	No	2.6, 1.1
No	Yes	2.4, 1.1
No	No	3.6, .73

Discussion

Whether or not communication was computer-mediated affected gesture production. Participants gestured more frequently in the unmediated settings. In the unmediated settings, seeing each other seemingly only increases gesture production. Yet in the mediated setting with mutual visibility, two factors may act in opposite directions. Our previously discussed results showed that in the mediated setting, being seen by the addressee increases gesture production, whereas seeing the addressee decreases gesture production. This may explain why participants gestured less frequently in the mediated setting. However, we did not find a difference in perceived interest of the addressee between the mediated and unmediated setting with mutual visibility.

Another possible explanation is a difference in affordances between mediated and unmediated communication (Brennan & Ohaeri, 1999). Even though one of the mediated settings offered live audio and video, narrators produced fewer gestures than in a face-to-face setting. The most notable difference between these two settings may be that the mediated setting did not enable interlocutors to look each other in the eyes. One either looks at the camera, or at the eyes of the other person, such that mutual gaze never occurs. We intend to address this factor in a follow-up study, by using a mediated setting that does allow for mutual gaze. Other factors such as not sharing a physical space may also be of influence, especially for pointing gestures (Isaacs & Tang, 2003).

General Discussion and Conclusion

Although our results suggest that several factors interact in our mediated settings, we found a clear effect of whether the addressee could see the speaker. Speakers produced representational hand gestures more frequently when they could be seen by their addressee, rather than when they could see their addressee, suggesting that speakers adjusted their gesturing to the addressee's perspective correctly. This is not to say that they never make mistakes in taking into account what their addressee can and cannot see during language production. Yet our results cannot be explained by assuming that speakers predominantly base their gesture production on their own visual perspective. Rather, they apply their knowledge of what the addressee can see correctly to their hand gesture production.

Acknowledgements

We like to thank Nelianne van den Berg for her help in collecting and coding the data. We also thank all participants as well as the anonymous reviewers.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? the costs and benefits of hedging. *SIGSOFT Softw. Eng. Notes, 24*(2), 227-235.

Cohen, A. A., & Harison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Language and Social Psychology, 8*, 211-288.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Personality and Social Psychology, 8*, 211-228.

Isaacs, E. A., & Tang, J. C. (2003). *What video can and can't do for collaboration: a case study.* Paper presented at the Multimedia, Anaheim, CA.

Jacobs, N., & Garnham, A. (2006). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 26*, 291-303.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396-414.

McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought*. Chicago and London: The University of Chicago Press.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-computer interactions. *Gesture, 9*(1), 97-126.

Wardow Lane, L., Groisman, M., & Ferreira, V., S. (2006). Don't talk about pink elephants: speakers' control over leaking private information during language production. *Psychological Science, 17*(4), 273-277.